·论著·

基于加权基因共表达网络和癌症基因组图谱临床数据分析并鉴定肝细胞癌的 Hub 基因研究

陈超[1]，陈天翔[2]，刘钱伟[1]，张秩[1]，王欢欢[2]，吴平平[1]，高磊[1]，于照祥[1*]

1.710077 陕西省西安市，西安医学院第一附属医院普通外科
2.710061 陕西省西安市，西安交通大学第一附属医院肝胆外科
*通信作者：于照祥，教授/主任医师；E-mail：yuzhaoxiang2@163.com

【摘要】 **背景** 肝细胞癌（HCC）是全球常见的癌症相关死亡的第三大原因，约占所有原发性肝癌病例的90%，其复发率和死亡率较高，目前发生的分子机制仍不清楚。**目的** 探索 HCC 潜在的分子机制，发掘新的生物标志物。**方法** 从 GEO 数据库中下载基因表达谱 GSE62232，从 TCGA 数据库下载 RNA-seq 表达数据和临床相关信息，通过差异基因表达分析正常肝脏组织与 HCC 组织的差异基因；对差异表达基因进行富集分析；基于 TCGA 中 HCC 和 GSE62232 的基因表达数据概况，使用 WGCNA R 包建立共表达网络，进行加权基因共表达网络分析（WGCNA），选择具有临床意义的模块，并筛选候选 Hub 基因；进一步分析候选 Hub 基因在 HCC 组织和正常肝脏组织显著差异表达、与 HCC 患者总体生存期和无病生存期是否显著相关，最终确定 Hub 基因；通过人类蛋白质图谱数据库对 Hub 基因蛋白表达进行验证。**结果** 本研究的基因表达数据来自 50 个正常肝脏组织样本和 373 个 HCC 组织样本。通过差异基因表达分析发现 7 230 个在 HCC 和正常肝脏组织之间差异表达的基因（HCC 中 3 691 个上调基因和 3 539 个下调基因）。富集分析表明，上调的差异表达基因主要参与细胞周期调控和有丝分裂过程；下调的差异表达基因主要参与小分子代谢和有机酸代谢等过程。WGCNA 确定了 19 个与 HCC 患者临床特征相关基因模块，通过分析模块与临床特征之间的关系，筛选出青色模块和紫色模块。青色模块基因中同时与患者总生存期和无病生存期强烈相关的前两个基因为 VPS45 和 FAM189B；紫色模块基因中同时与患者总生存期和无病生存期强烈相关的前两个基因分别为 CLEC1B 和 FCN3，因此将 VPS45、FAM189B、CLEC1B 和 FCN3 确定为最终的 Hub 基因。人类蛋白质图谱数据库免疫组织化学染色显示：VPS45 和 FAM189B 在 HCC 组织中的表达高于正常肝脏组织，FCN3 在 HCC 组织中的表达低于正常肝脏组织，CLEC1B 在 HCC 组织和正常肝脏组织中表达差异不明显。**结论** 初步确定 VPS45、FAM189B、CLEC1B 和 FCN3 可能是 HCC 的新型潜在生物标志物，这些 Hub 基因可能为 HCC 的靶向治疗提供理论基础。

【关键词】 肝细胞癌；加权基因共表达网络分析；Hub 基因；分子靶向治疗；VPS45；FAM189B；CLEC1B；FCN3

【中图分类号】 R 730.261 【文献标识码】 A DOI：10.12114/j.issn.1007-9572.2023.0243

Analysis and Identification of Hub Genes in Hepatocellular Carcinoma Based on Weighted Gene Co-expression Network and Cancer Genome Atlas Clinical Data

CHEN Chao[1]，CHEN Tianxiang[2]，LIU Qianwei[1]，ZHANG Zhi[1]，WANG Huanhuan[2]，WU Pingping[1]，GAO Lei[1]，YU Zhaoxiang[1]*

1.Department of General Surgery，the First Affiliated Hospital of Xi'an Medical University，Xi'an 710077，China
2.Department of Hepatobiliary Surgery，the First Affiliated Hospital of Xi'an Jiaotong University，Xi'an 710061，China
**Corresponding author：YU Zhaoxiag，Professor/Chief physician；E-mail：yuzhaoxiang2@163.com*

【Abstract】 **Background** Hepatocellular carcinoma（HCC） is the third leading cause of common cancer-related

基金项目：陕西省科技厅面上项目（2021JM-503）
引用本文：陈超，陈天翔，刘钱伟，等.基于加权基因共表达网络和癌症基因组图谱临床数据分析并鉴定肝细胞癌的 Hub 基因研究［J］.中国全科医学，2023.［Epub ahead of print］.DOI：10.12114/j.issn.1007-9572.2023.0243.［www.chinagp.net］
CHEN C，CHEN T X，LIU Q W，et al. Analysis and identification of Hub genes in hepatocellular carcinoma based on weighted gene co-expression network and cancer genome atlas clinical data［J］. Chinese General Practice，2023.［Epub ahead of print］.

mortality globally, accounting for approximately 90% of all primary liver cancer cases. Its recurrence and mortality rates are high, with the underlying molecular mechanisms remaining unclear. **Objective** To explore potential molecular mechanisms of HCC and explore novel biomarkers. **Methods** Gene expression profile GSE62232 was retrieved from the GEO database, RNA-seq expression data and clinical information were retrieved from TCGA database, differential gene expression analysis was conducted between normal liver tissue and HCC tissue. Enrichment analysis on the differentially expressed genes was performed. Based on the gene expression data profiles of HCC and GSE62232 in TCGA, a co-expression network was established using the WGCNA R package, and weighted gene co-expression network analysis (WGCNA) was performed to select clinically significant modules and screen candidate Hub genes; the candidate Hub genes were further analyzed for significant differential expression in HCC tissues and normal liver tissues, and whether they were significantly correlated with the overall survival and disease-free survival of HCC patients. The hub genes were conclusively identified, and their protein expression was validated through the Human Protein Atlas database. **Results** The genetic expression data in this study were obtained from 50 normal liver tissue samples and 373 HCC tissue samples. Through differential gene expression analysis, a total of 7 230 genes differential expression between HCC and normal hepatic tissue, comprising 3 691 up-regulated genes and 3 539 down-regulated genes in HCC were identified. Enrichment analysis showed that the up-regulated differentially expressed genes were mainly involved in cell cycle regulation and mitotic processes; the down-regulated differentially expressed genes were mainly involved in processes such as small molecule metabolism and organic acid metabolism. WGCNA identified 19 gene modules related to the clinical features of HCC patients, the cyan and purple modules were screened by analyzing the relationship between the modules and the clinical features. The first two genes in the cyan module genes that were strongly associated with both overall survival and disease-free survival of patients were VPS45 and FAM189B. In the purple module genes, first two genes that were strongly associated with both overall survival and disease-free survival of patients were CLEC1B and FCN3, respectively; therefore, VPS45, FAM189B, CLEC1B and FCN3 were identified as the final Hub genes. Immunohistochemical staining in the Human Protein Atlas database showed that VPS45 and FAM189B were expressed higher in HCC tissues than in normal liver tissues. FCN3 was expressed in HCC tissues lower than in normal liver tissues, the difference in the expression of CLEC1B between HCC tissues and normal liver tissues was not obvious. **Conclusion** VPS45, FAM189B, CLEC1B and FCN3 have been preliminary identified as possible novel potential biomarkers for HCC, which may provide a theoretical basis for targeted therapy of HCC.

【Key words】 Hepatocellular carcinoma; Weighted gene co-expression network analysis; Hub gene; Molecular Targeted Therapy; VPS45; FAM189B; CLEC1B; FCN3

肝细胞癌（HCC）是癌症中相关死亡的第三大原因[1]，也是中国近年来癌症相关死亡的主要原因。HCC 在所有原发性肝癌病例中占 90%[2]。HCC 发病机制的主要危险因素包括慢性乙型肝炎病毒（HBV）和丙型肝炎病毒（HCV）感染、吸烟、饮酒、超重和非酒精性脂肪性肝病（NAFLD）、糖尿病以及黄曲霉毒素 B1 摄入量等[3]，这些因素导致 DNA 损伤，表观遗传改变和癌症相关突变，最终导致 HCC 进展[4]。绝大多数患者确诊时已是晚期。尽管 HCC 的治疗在近几年取得很大进展，主要包括经导管动脉化疗栓塞、分子靶向治疗、消融治疗、手术切除和肝移植等[3, 5]，但 HCC 患者的 5 年生存率仍然很低[6]，并且其 5 年复发率高达 80%~90%，预后较差[7]。HCC 发生和发展的确切机制仍不清楚，因此，阐明 HCC 发生、发展的分子机制，开发新的诊断和治疗方法来改善 HCC 的临床结果是非常迫切和必要的。

近年来，随着高通量测序技术不断发展，为癌症的基因组学、转录组学和表观基因组学的特征提供了新的研究方法。表达谱的生物信息学分析已被广泛用于鉴定新的和更有效的潜在生物标志物，用于癌症的治疗和患者预后[8]。通过生物信息学的大数据分析，可以有效整合复杂疾病，尤其是癌症的多个大规模数据集[9]。加权基因共表达网络分析（weighted gene co-expression network analysis，WGCNA）是一种先进的方法，其具有准确和高效的广泛基因分析优点，用于基于相似的基因表达模式构建共表达模块，并分析临床特征模块与不同基因组之间的相关性[10]。WGCNA 已被广泛用于识别不同类型癌症相关的临床特征模块和 Hub 基因。如一项基于 WGCNA 研究将 6 个 Hub 基因和人肾细胞癌的进展以及患者的预后联系起来[11]，鉴定了在侵袭性肾上腺皮质恶性肿瘤中高表达并与总生存期呈明显负相关的 4 个 Hub 基因（TOP2A、CHEK1、TTK 和 CENPA）[12]。

本研究通过差异基因表达分析和 WGCNA 对从 TCGA 和 GEO 数据库中 HCC 的 mRNA 数据，在表达和功能水平上进行分析。然后进行功能富集分析，确定与 HCC 患者临床特征相关的模块，以了解这些共表达基因的潜在生物学功能。通过这些生物信息学分析鉴定出

Hub 基因，并结合生存分析、HCC 组织和正常肝脏组织表达差异分析、免疫组织化学染色分析和文献分析进行验证。这些结果将有助于临床了解 HCC 的病因和潜在的分子机制，并为 HCC 提供新的治疗靶点或生物标志物。

1 资料与方法

1.1 数据来源和数据处理

GSE62232［HCC 对应 81 例患者。在所有病例中，诊断时肝切除后肿瘤样本被冷冻（-80 ℃），使用正常肝脏样本作为参考样本。使用 Affymetrix U133plus v2 阵列（GPL570）进行比较基因表达分析］基因表达谱来自 GEO 数据库（http：//www.ncbi.nlm.nih.gov/geo/），基于 GPL570 平台（［HG-U133_Plus_2］Affymetrix Human Genome U133 Plus 2.0 Array）。从 TCGA 数据库[13]（https：//www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga）中下载 HCC 患者的 RNA 测序数据（50 个正常样本和 373 个 HCC 样本）、临床信息和生存信息（59 个正常样本和 379 个 HCC 样本），患者临床信息数据和样本量见表 1。根据注释文档将探针转化为基因符号，通过测定所有对应探针的中位表达值，去除同一基因的重复探针。根据处理的结果，总共有 11 627 个基因被选中用于后续分析。

表 1 TCGA 数据库的临床信息和样本
Table 1 Clinical information and samples from the TCGA database

特征	合计（N=438）	生存（N=266）	死亡（N=172）
年龄（$\bar{x}\pm s$，岁）	59.98±13.82	58.58±13.49	62.14±14.09
年龄［M（MIN，MAX）］	62.00［16.00，90.00］	61.00［16.00，84.00］	65.00［18.00，90.00］
性别			
女	146（33.33）	79（18.04）	67（15.30）
男	292（66.67）	187（42.69）	105（23.97）
微血管侵犯[a]			
否	248（66.13）	162（43.20）	86（22.93）
是	127（33.87）	81（21.60）	46（12.27）
体质量（$\bar{x}\pm s$，kg）	73.40±19.64	73.68±20.93	72.96±17.46
体质量［M（MIN，MAX）］	70.00［40.00，172.00］	69.00［40.00，172.00］	73.00［40.00，139.00］
BMI>24 kg/m^2[a]			
否	185（47.19）	116（29.59）	69（17.60）
是	207（52.81）	132（33.67）	75（19.13）
TNM 分期（期）[a]			
Ⅰ	198（49.01）	141（34.90）	57（14.11）
Ⅱ	100（24.75）	66（16.34）	34（8.42）
Ⅲ	100（24.75）	46（11.39）	54（13.37）
Ⅳ	6（1.49）	2（0.50）	4（0.99）

注：[a] 表示数据有缺失。

1.2 HCC 中差异表达基因的筛选

对于降维方法 UMAP，首先使用 R 软件包 UMAP（version 0.2.7.0）进行分析，对表达谱进行 z-score，再使用 UMAP 函数进行降维分析以获得降维后的矩阵。Limma（linear models for microarray data）[14]是一种基于广义线性模型的差异表达筛选方法，用于识别正常肝脏组织和 HCC 组织之间的差异表达基因，在此处使用 R 软件包 limma（version 3.40.6）进行差异分析，以获得正常肝脏组织与 HCC 组织的差异基因。获取的表达谱数据集，利用 lmFit 函数进行多元线性回归分析，进一步使用 eBays 函数进行分析，最终获得显著性差异的基因。

1.3 差异表达基因的功能富集

对于基因集进行功能富集分析，使用 R 软件包 org.Hs.eg.db（version 3.1.0）中的基因 GO 注释，将其作为背景把基因映射到背景集合中，使用 R 软件包 clusterProfiler（version 3.14.3）进行富集分析，从而获取基因集富集的结果。设定最小基因集为 5，最大基因集为 5 000，P<0.05 和 FDR<0.1 作为显著性差异。

1.4 加权基因共表达网络分析

加权基因共表达网络分析是一种分析多个样本基因表达模式的分析方法，其可以将具有相似表达模式的基因聚类，并分析模块与特定性状或表型之间的关系。基于 TCGA 中 HCC 和 GSE62232 的基因表达数据概况，使用 WFCNA R 包建立共表达网络。以基因表达谱为例：首先，利用基因表达谱，去除了在各个样本中标准差为 0 的基因，利用 R 软件包 WGCNA 的 goodSamplesGenes 方法去除离群的基因以及样本，进一步使用 WGCNA 构建一个无标度共表达网络。具体而言，第一步先对所有成对基因进行皮尔逊相关矩阵和平均连锁法，然后使用幂函数 $a_m n=|C_{mn}|$^β（$C_m n$=Gene_m 和 Gene_n）之间的 Pearson′s 相关；构建出加权邻接矩阵。β 是一个软阈值参数，其可以强调基因之间的强相关性。在选择 5 的幂之后，将邻接变换为拓扑重叠矩阵（Topological Overlap Matrix，TOM）TOMi，j=（lij+aij）/（min（ki+kj）+1-aij），该矩阵可以测量一个基因的网络连通性，该网络连通性定义为其与所有其他基因的邻接之和，用于网络基因定量，并计算相应的相异性。为了将具有相似表达谱的基因分类为基因模块，根据基于 TOM 的不相似性度量进行平均连锁分级聚类，基因树图的最小大小（基因组）为 50。此外还合并了距离 <0.25 的模块，最终获得了 19 个共表达模块，其中 grey 模块被认为是无法被分配给任何模块的基因集合。

1.5 具有临床意义的模块的选择和 HCC 中的 Hub 基因的鉴定

首先，主成分分析用于描述模块特征基因，对应于

每个模块内所有基因的单一特征表达谱。计算临床特征和这些特征基因之间的相关性，以确定哪些模块具有临床意义。基因表达和临床特征之间的线性关系被赋予一个基因显着性，等于单个基因 P 值的对数。如果基因显着性与模块成员密切相关，定义为模块的特征基因与个体基因表达谱之间的相关性，从而得出结论该模块的中心基因与 HCC[15]相关，将这些基因视为候选 Hub 基因。如先前研究[16]所示：计算与基因的表达相关性以获得基因显著性，同时计算模块特征向量与基因的表达相关性以获得模块成员，根据截止标准，76 个在临床显著模块中具有高连接性的基因被鉴定为候选 Hub 基因。

1.6 Hub 基因的基因验证和生物信息学验证

使用 GEPIA 在线网站（http：//gepia.cancer-pku.cn/）分析 HCC 样本中 Hub 基因的表达水平，并基于 Kaplan-Meier 分析使用 R 套件中的生存包（版本：3.2-7）。首先，从 TCGA 获得 359 个 HCC 肿瘤样本的差异基因表达谱和预后数据，然后确定每个基因的中位数表达值，根据给定基因的表达水平是高于还是低于中位数，样本被分配到给定基因的“高表达”或“低表达”组。使用对数秩检验评估高表达组或低表达组之间的总体生存期和无病生存期的显著性。如果 $P<0.05$，认为该基因是经过验证的 Hub 基因。然后，根据来自 TCGA 和 GEPIA 网站上的数据，筛选正常肝脏组织和 HCC 组织之间 Hub 基因表达的差异。表达水平通过平均值归一化，并且与 $P<0.05$ 相关的差异被认为具有统计学意义。

1.7 人类蛋白质图谱数据库对 Hub 基因蛋白表达的验证

人类蛋白质图谱数据库（https：//www.proteinatlas.org/）主要用于为提供各种人类蛋白质的组织和细胞分布信息。使用来自人类蛋白质图谱数据库的免疫组织化学染色结果验证了 HCC 组织和正常肝脏组织之间生存相关基因的蛋白质表达。

2 结果

2.1 数据预处理

本研究的表达数据来自 50 个正常肝脏组织样本和 373 个 HCC 组织样本。基于主成分分析从数据中排除了 14 个肿瘤样本（图 1A）。最终来自这 409 个样本的基因表达谱用于后续分析。

2.2 HCC 样本中差异表达基因的鉴定及 GO 富集分析

在 50 个正常样本和 359 个 HCC 样本之间共鉴定出 7 230 个差异表达基因，差异表达基因火山图见图 1B，包括 3 691 个上调基因和 3 539 个下调基因，差异分析结果见表 2。为了探索差异表达基因在 HCC 中的潜在生物学功能，本研究对其进行 GO 富集分析。

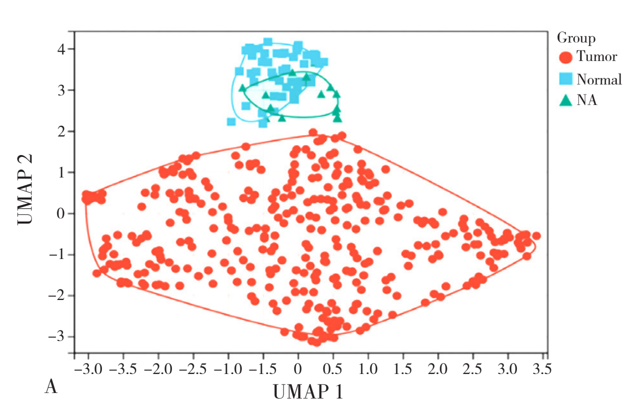
上调的差异表达基因主要参与细胞周期调控、有丝分裂过程、核分裂和染色体分离等（图 2A、2C）；相反，下调的差异表达基因主要参与对外刺激反应、小分子代谢和有机酸代谢等过程（图 2B、2D）。

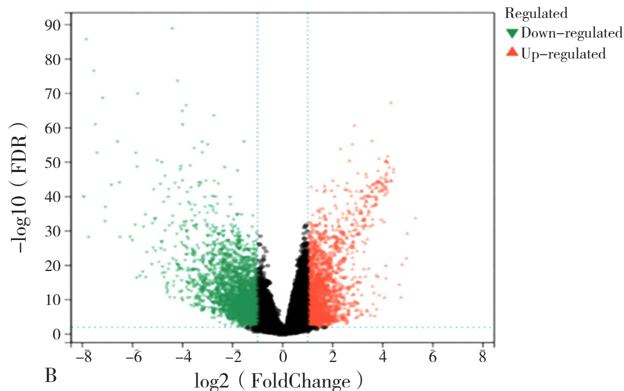
比较 HCC 组织和正常肝脏组织的基因表达差异，获得表达数据集，根据差异倍数为 1.5 倍，$P<0.05$ 为显著差异，进一步利用 lmFit 函数进行多元线性回归分析，再使用 eBays 函数进行分析，最终获得每个基因的差异显著性，并绘制差异表达基因热图。

表 2 差异分析结果统计摘要
Table 2 Statistical summary of variance analysis results

显著性阈值	1.2 倍差异		1.3 倍差异		1.5 倍差异		2 倍差异	
	上调	下调	上调	下调	上调	下调	上调	下调
$P<0.05$	6 570	5 070	5 479	4 491	3 691	3 539	1 622	2 063
$P<0.01$	5 942	4 591	5 109	4 175	3 594	3 421	1 613	2 057
FDR<0.05	6 384	4 912	5 370	4 392	3 667	3 502	1 621	2 061
FDR<0.01	5 687	4 380	4 953	4 022	3 526	3 342	1 604	2 049

注：A 和 C 为上调基因功能富集 GO 分析的气泡图和圈图；B 和 D 为下调基因功能富集 GO 分析的气泡图和圈图。





注：A 为主成分分析：红色点代表肿瘤组织，蓝色点代表正常组织，绿色点代表排除的 HCC 样本；B 为火山图：绿色点代表 HCC 中下调的基因，黑色点代表 HCC 和正常肝脏组织之间没有显著差异的基因，红色点代表 HCC 中上调的基因。

图 1 鉴定 50 个正常肝脏组织样本和 359 个 HCC 组织样本之间的差异表达基因
Figure 1 Differentially expressed genes were identified between 50 normal samples and 359 HCC samples

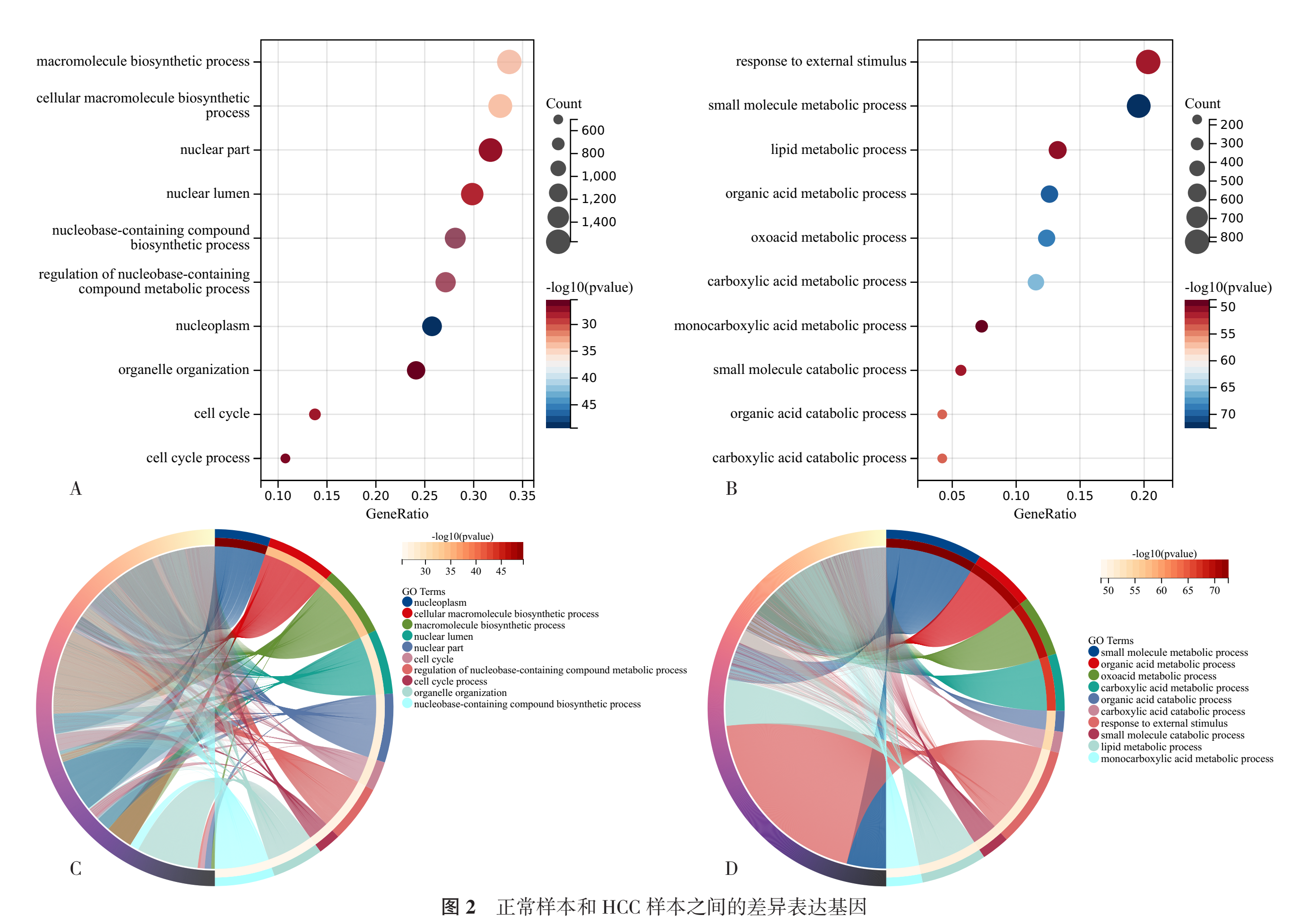


图 2 正常样本和 HCC 样本之间的差异表达基因

Figure 2 Common differentially expressed genes were identified between normal and HCC samples

2.3 WGCNA 和关键模块的识别

加权基因共表达网络分析基于 7 230 个差异表达基因的表达矩阵和 409 个 HCC 样本的临床数据。进行聚类分析检查 409 个样本的数据质量，结果显示所有样本在聚类中并且在截止阈值内，加权基因共表达网络分析中应用了 6 个临床变量：疾病状态（Tumor or Normal）、性别、年龄、体质量、TNM 分期和微血管侵犯（图 3A）。409 个样本分为两个簇，肿瘤和正常，HCC 样本的临床特征和数据的聚类树状图（图 3B）。

为了构建无标度网络，将软阈值 β 设置为 5，独立度设置为 0.86，平均连通性接近 0（图 4A、4B）。模块参数设置为最小模块 30，敏感性为 3，模块合并阈值为 0.25，将其中具有相似表达模式的差异表达基因聚集到相同的模块中，此外合并距离小于 0.25 的模块，最终获得了 19 个共表达模块，并绘制模块特征基因热图（图 4C），其中 grey 模块是无法被分配给任何模块的基因集合。

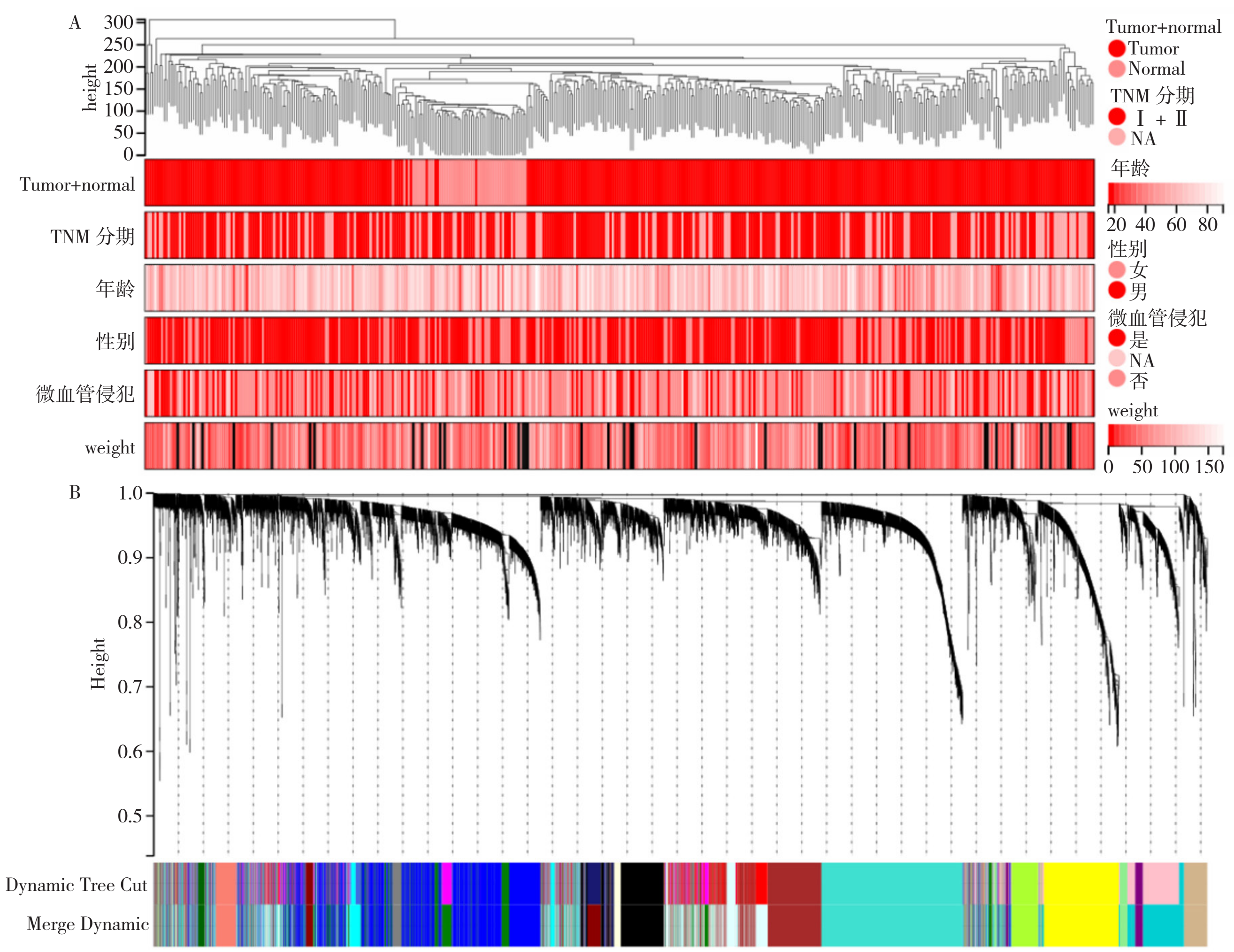
然后尝试评估模块与临床特征之间的关系，发现青色模块的特征基因与 HCC 之间具有强烈的正相关（cor=0.64，$P=3.6\times10^{-50}$），而紫色模块的特征基因与 HCC 之间具有高度负相关（cor=−0.80，$P=5.6\times10^{-97}$）。结果表明，青色模块可能促进 HCC 的肿瘤发生，而紫色模块可能会预防 HCC。因此，分析青色模块和紫色模块的 Hub 基因。

2.4 从青色和紫色模块中识别候选 Hub 基因

MM 和 GS 分数在青色和紫色模块中彼此呈正相关（图 5）。青色模块中选择 Hub 基因的标准相对低于标准截止阈值（MM>0.8）。在青色模块中，选择满足“cor.gene Module Membership”>0.7 和“cor.gene Trait Significance”>0.5 阈值的前 8 个基因，分别为 TOMM40L、VPS45、MSTO1、FAM189B、TTC13、PYGO2、NVL 和 EHMT2。在紫色模块中，选择满足“cor.gene Module Membership”>0.8 和“cor.gene Trait Significance”>0.7 阈值的前 16 个基因，分别为 CLEC4M、BMP10、CLEC1B、CLEC4G、GDF2、NDST3、BMPER、STAB2、CCL23、CHRM2、COL6A6、CRHBP、FCN3 和 CCBE1。

2.5 Hub 基因表达及其与生存的相关性

基于 GEPIA2 在线数据库中肿瘤样本的表达数据和临床信息，分析青色模块中识别的基因和紫色模块中识



注：A 为加权基因共表达网络分析，红色代表变量“Tumor-Normal”的“肿瘤”，变量“体质量”，变量“性别”的“女性”或“男性”，变量“TNM 分期”以及“微血管侵犯”，红色的颜色程度与值呈正比；B 为聚类树状图。

图 3 HCC 样本的临床特征和数据的基因聚类图

Figure 3 Genetic clustering maps of clinical features and data of HCC samples

别的基因的表达与患者生存之间的潜在关联（图 6）。青色模块基因中仅与患者总生存期强烈相关的前 3 个基因分别为 TOMM40L、VPS45 和 FAM189B；青色模块基因中同时与患者总生存期和无病生存期强烈相关的前两个基因为 VPS45 和 FAM189B。紫色模块基因中仅与患者总生存期强烈相关的前 3 个基因分别为 CLEC1B、CCL23 和 FCN3；紫色模块基因中同时与患者总生存期和无病生存期强烈相关的前两个基因分别为 CLEC1B 和 FCN3。因此，将 VPS45、FAM189B、CLEC1B 和 FCN3 确定为最终的 Hub 基因。两个模块中与无病生存期有显著差异的基因有 VPS45、FAM189B、CLEC1B、FCN3 和 BMPER。

使用 GEPIA 网站，分析这些模块基因在 HCC 组织和正常组织之间表达情况（图 7），结果显示，VPS45 和 FAM189B 在 HCC 组织中上调，而 CLEC1B 和 FCN3 在 HCC 组织中下调。使用 GEO 数据库的数据获得了类似的结果。

2.6 免疫组织化学染色验证

从人类蛋白质图谱（The Human Protein Atlas，HPA）数据库[17]（网址：Human Protein Atlas proteinatlas.org）中获得的免疫组织化学染色显示：VPS45 和 FAM189B 在 HCC 组织中的表达高于正常肝脏组织，而 FCN3 在 HCC 组织中的表达低于正常肝脏组织，CLEC1B 在 HCC 组织和正常肝脏组织中表达差异不明显（图 8）。

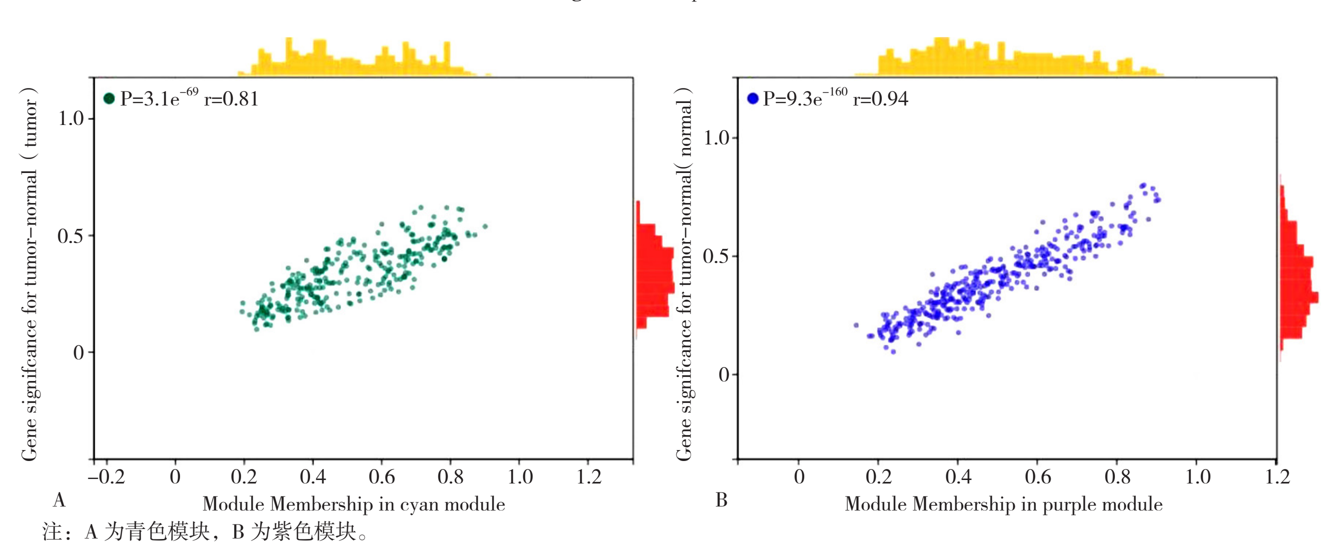
3 讨论

HCC 筛查指对符合 HCC 高风险的患者定期进行检查，其目标是在早期时间检测出 HCC 并予以及时干预，使患者存活率和生存质量提高。在 HCC 的临床应用中，常见的传统筛查方式包括肝脏超声检查和血清甲胎蛋白，其敏感性和特异性相对有限。因此，开发出新的更

注：A 为标度独立性；B 为平均连通性；C 为模块特征向量聚类。

图 4 共表达模块

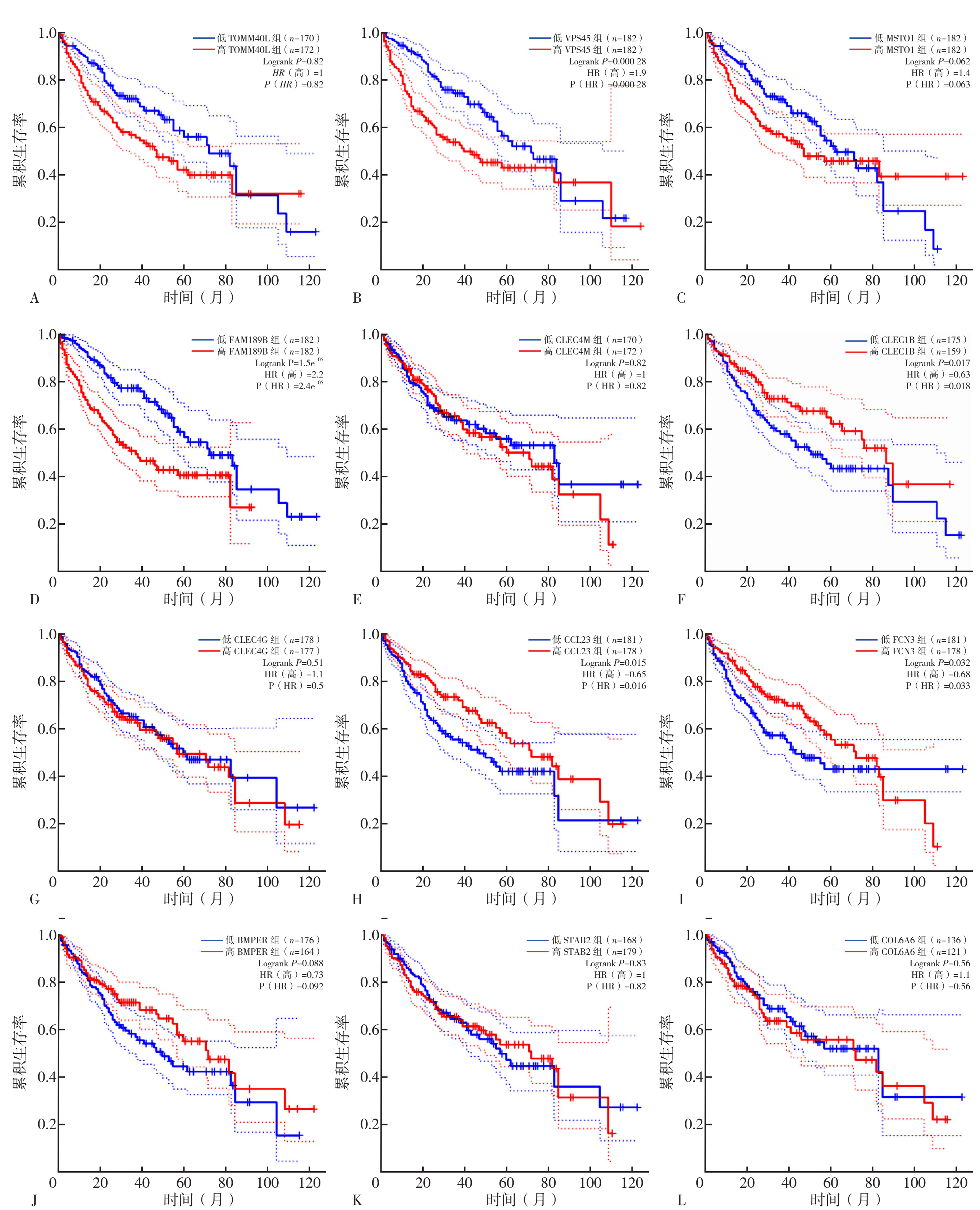
Figure 4 Coexpression module



注：A 为青色模块，B 为紫色模块。

图 5 模块特征和模块基因的相关性

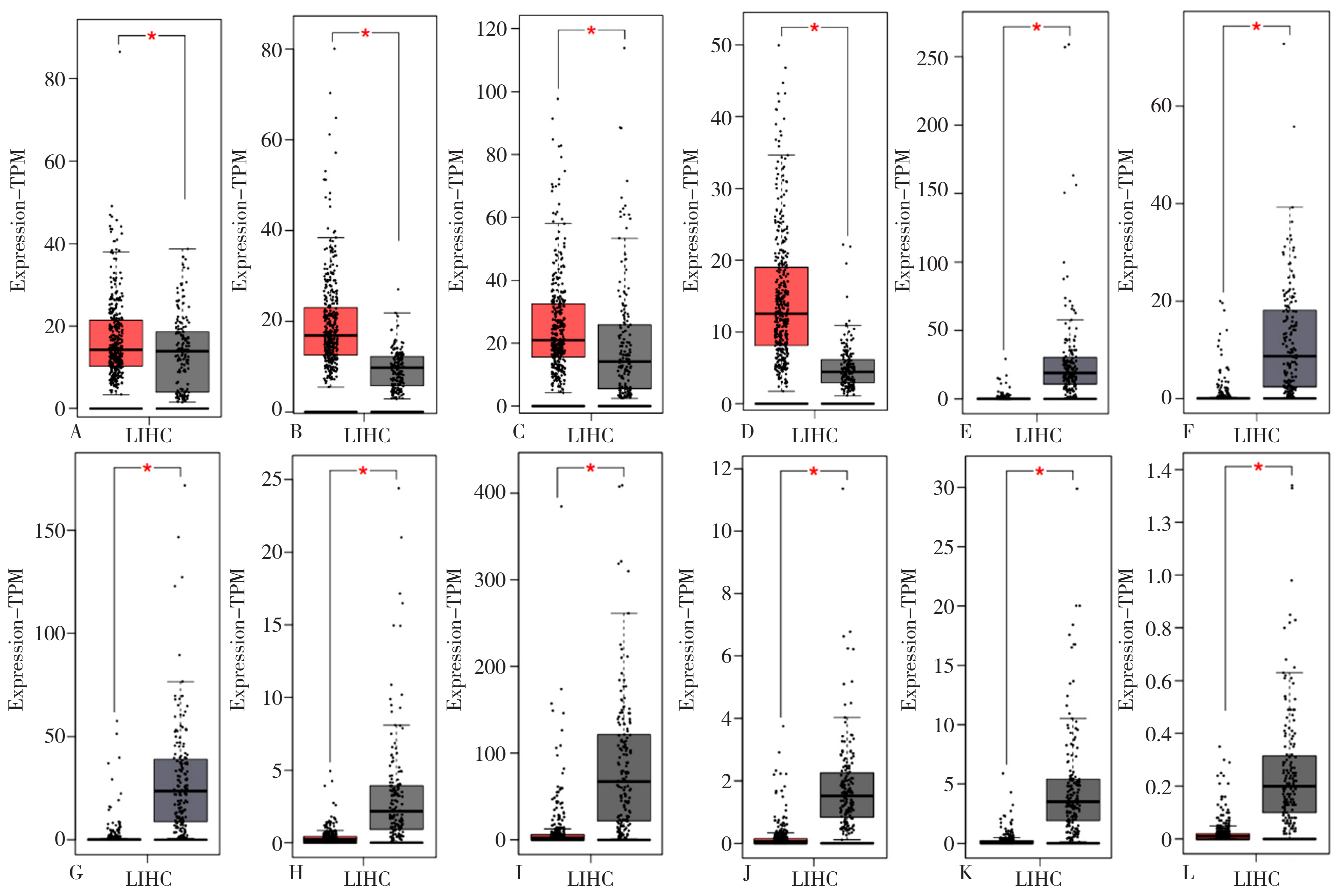
Figure 5 Correlation between module characteristics and module genes



注：A 为 TOMM40L，B 为 VPS45，C 为 MSTO1，D 为 FAM189B，E 为 CLEC4M，F 为 CLEC1B，G 为 CLEC4G，H 为 CCL23，I 为 FCN3，J 为 BMPER，K 为 STAB2，L 为 COL6A6；红线代表基因表达高的样本，蓝线代表基因表达低的样本。

图 6　HCC 组织中青色和紫色模块基因表达与患者总生存期的关系

Figure 6　Relationship between cyan and purple module gene expression and overall survival of patients with HCC



注：A 为 TOMM40L，B 为 VPS45，C 为 MSTO1，D 为 FAM189B，E 为 CLEC4M，F 为 CLEC1B，G 为 CLEC4G，H 为 CCL23，I 为 FCN3，J 为 BMPER，K 为 STAB2，L 为 COL6A6；$^{*}P$<0.05。

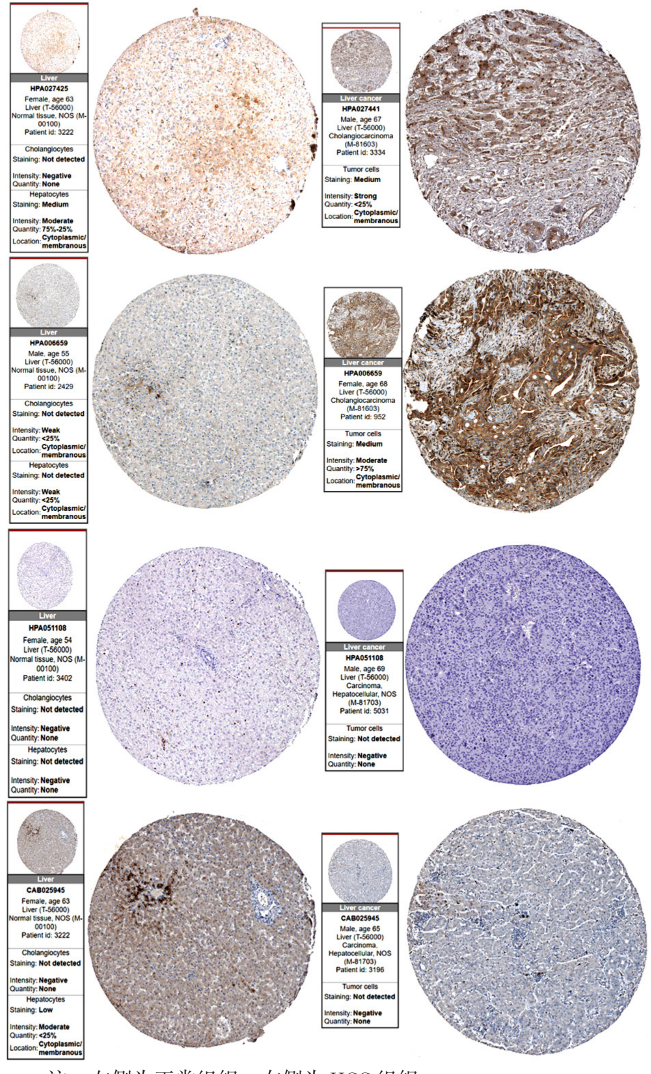
图 7 TCGA 数据库中青色和紫色模块基因在 HCC 组织及正常肝组织中表达水平

Figure 7 Expression levels of cyan and purple module genes in HCC tissues and normal liver tissues in TCGA database

准确的检测方式十分具有临床应用价值。目前研究表明几个基因已被确定为 HCC 诊断的新型生物标志物。例如，长链非编码 RNA KDM4A-AS1 在许多肿瘤中过度表达，尤其是在 HCC 中，长链非编码 RNA KDM4A-AS1 水平与疾病分期和肿瘤分级呈正相关，与患者生存呈负相关[18]，LV 等[19]研究显示 BAI3 和 CKAP2L 可能是结直肠癌的潜在预后因子和治疗靶点。

在这项研究中，将差异表达基因和加权基因共表达网络结合起来，以提高识别 HCC 相关基因的能力。差异基因功能富集显示，主要参与细胞周期调控的基因在 HCC 组织中失调，与近些年相关研究结果一致[20]。正常细胞周期的紊乱是导致癌症的原因之一，靶向调节癌细胞周期是一种潜在的治疗方法[21]。生物信息学分析基于 TCGA 和 GEO 数据库，分析在 HCC 中重要模块和 Hub 基因，确定了两个关键模块，分别为青色模块和紫色模块。分析这两个模块基因与患者总生存期和无病生存期的相关性，选择了青色模块中显著相关的前两个基因 VPS45 和 FAM189B，选择了紫色模块中显著相关的前两个基因 CLEC1B 和 FCN3，这 4 个基因作为入选的 Hub 基因。

VPS45 缺陷型中性粒细胞和成纤维细胞表面的 β1 整合素水平降低，VPS45 缺陷型成纤维细胞运动能力受损和细胞凋亡增加[22]。另外有研究显示 VPS45 的表达水平和多种癌症相关，比如 YAMANOI 等[23]研究表明 VPS45 表达水平和恶性肿瘤卵巢癌相关。FAM189B 又称为 COTE1，通过机制分析表明，FAM189B 可以与肿瘤抑制因子结构域氧化还原酶发生物理关联，FAM189B 与 HCC 细胞的侵袭密切相关[24]；FAM189B 蛋白和 mRNA 的过表达可能会增加胃癌的发生率，并且细胞周期的通路是其 KEGG 富集分析最具有显著差异的通路[25]。CLEC1B 是 C 型凝集素结构域家族 1 成员，并且其与肝细胞癌的免疫浸润相关，过表达的 CLEC1B 抑制 HuH7 细胞的增殖和迁移能力[26]。FCN3 是一种分泌型凝集素，能够激活补体通路，FCN3 的异位表达能够激活内质网应激未折叠蛋白，抑制内质网应激反应可提高肺腺癌细胞的存活率，FCN 通过诱导内质网应激发挥抑癌作用[27]。本研究方法具有数据挖掘的局限性，研究所用数据来源于网络数据库，数据样本可能存



注：左侧为正常组织，右侧为 HCC 组织。

图 8 模块基因的免疫组织化学染色

Figure 8 Immunohistochemical staining of modular genes

在偏倚，测序结果在技术上可能存在偏差，为了提高 WGCNA 结果的可靠性，使用人类蛋白质图谱的免疫组织化学数据进行确认，然而由于数据库的限制，无法获得每个基因的肿瘤和邻近正常样本的所有相关 IHC 数据。获得的 Hub 基因在肝癌预后中的分子调节机制还需要进一步通过临床数据和基础实验来证实，其可能作为肝癌早期筛查分子标志物、预后分子标志物及肝癌治疗靶点，为人群筛查或肝癌患者治疗提供帮助。

总之，本研究进行了全面的生物信息学分析，以确定 HCC 和正常肝脏组织之间的潜在生物预测标志物，本研究结果表明，VPS45、FAM189B、CLEC1B 和 FCN3 可能是 HCC 的新型潜在生物标志物，具有特殊临床意义，然而需要在大量临床样本中得到真实验证，未来进行更进一步探索。

作者贡献：陈超提出研究目标，负责数据分析和写作；陈天翔负责数据验证；刘钱伟、张秩、王欢欢、高磊和吴平平负责数据整理；于照祥负责文章的质量控制与审查，对文章整体负责，监督管理。

本文无利益冲突。

参考文献

[1] SUNG H，FERLAY J，SIEGEL R L，et al. Global cancer statistics 2020：GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin，2021，71(3)：209–249. DOI：10.3322/caac.21660.

[2] MEHTA A，HERRERA H，BLOCK T. Glycosylation and liver cancer[J]. Adv Cancer Res，2015，126：257–279. DOI：10.1016/bs.acr.2014.11.005.

[3] VOGEL A，MEYER T，SAPISOCHIN G，et al. Hepatocellular carcinoma[J]. Lancet，2022，400（10360）：1345–1362. DOI：10.1016/s0140–6736（22）01200–4.

[4] Cancer Genome Atlas Research Network Electronic Address. Comprehensive and integrative genomic characterization of hepatocellular carcinoma[J]. Cell，2017，169（7）：1327–1341.e23. DOI：10.1016/j.cell.2017.05.046.

[5] LLOVET J M，REAL M I，MONTAÑA X，et al. Arterial embolisation or chemoembolisation versus symptomatic treatment in patients with unresectable hepatocellular carcinoma：a randomised controlled trial[J]. Lancet，2002，359（9319）：1734–1739. DOI：10.1016/S0140–6736（02）08649–X.

[6] SIEGEL R L，MILLER K D，JEMAL A. Cancer statistics，2017[J]. CA Cancer J Clin，2017，67（1）：7–30. DOI：10.3322/caac.21387.

[7] XIA F，WU L L，LAU W Y，et al. Adjuvant sorafenib after heptectomy for Barcelona clinic liver cancer–stage C hepatocellular carcinoma patients[J]. World J Gastroenterol，2016，22（23）：5384–5392. DOI：10.3748/wjg.v22.i23.5384.

[8] LI N，ZHAN X Q. Identification of clinical trait–related lncRNA and mRNA biomarkers with weighted gene co–expression network analysis as useful tool for personalized medicine in ovarian cancer[J]. EPMA J，2019，10（3）：273–290. DOI：10.1007/s13167–019–00175–0.

[9] KUENZI B M，IDEKER T. A census of pathway maps in cancer systems biology[J]. Nat Rev Cancer，2020，20（4）：233–246. DOI：10.1038/s41568–020–0240–7.

[10] LANGFELDER P，HORVATH S. WGCNA：an R package for weighted correlation network analysis[J]. BMC Bioinformatics，2008，9：559. DOI：10.1186/1471–2105–9–559.

[11] YUAN L S，CHEN L，QIAN K Y，et al. Co–expression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma（ccRCC）[J]. Genom Data，2017，14：132–140. DOI：10.1016/j.gdata.2017.10.006.

[12] XIA W X，YU Q，LI G H，et al. Identification of four hub genes associated with adrenocortical carcinoma progression by WGCNA[J]. PeerJ，2019，7：e6555. DOI：10.7717/peerj.6555.

[13] COLAPRICO A，SILVA T C，OLSEN C，et al. TCGAbiolinks：an R/Bioconductor package for integrative analysis of TCGA data[J]. Nucleic Acids Res，2016，44（8）：e71. DOI：10.1093/nar/

gkv1507.
[14] RITCHIE M E, PHIPSON B, WU D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies [J]. Nucleic Acids Res, 2015, 43 (7): e47. DOI: 10.1093/nar/gkv007.
[15] DI Y, CHEN D S, YU W, et al. Bladder cancer stage-associated hub genes revealed by WGCNA co-expression network analysis [J]. Hereditas, 2019, 156: 7. DOI: 10.1186/s41065-019-0083-y.
[16] TANG J N, KONG D G, CUI Q X, et al. Prognostic genes of breast cancer identified by gene co-expression network analysis [J]. Front Oncol, 2018, 8: 374. DOI: 10.3389/fonc.2018.00374.
[17] KARLSSON M, ZHANG C, MÉAR L, et al. A single-cell type transcriptomics map of human tissues [J]. Sci Adv, 2021, 7 (31): eabh2169. DOI: 10.1126/sciadv.abh2169.
[18] CHEN T X, LIU R K, NIU Y S, et al. HIF-1α-activated long non-coding RNA KDM4A-AS1 promotes hepatocellular carcinoma progression via the miR-411-5p/KPNA2/AKT pathway [J]. Cell Death Dis, 2021, 12 (12): 1152. DOI: 10.1038/s41419-021-04449-2.
[19] LV Y M, XIE B B, BAI B J, et al. Weighted gene coexpression analysis indicates that PLAGL2 and POFUT1 are related to the differential features of proximal and distal colorectal cancer [J]. Oncol Rep, 2019, 42 (6): 2473-2485. DOI: 10.3892/or.2019.7368.
[20] SCHAFER K. The cell cycle: a review [J]. Vet Pathol, 1998, 35: 461-478. DOI: 10.1177/030098589803500601.
[21] AARTS M, LINARDOPOULOS S, TURNER N C. Tumour selective targeting of cell cycle kinases for cancer treatment [J]. Curr Opin Pharmacol, 2013, 13 (4): 529-535. DOI: 10.1016/j.coph.2013.03.012.
[22] VILBOUX T, LEV A, MALICDAN M C V, et al. A congenital neutrophil defect syndrome associated with mutations in VPS45 [J]. N Engl J Med, 2013, 369 (1): 54-65. DOI: 10.1056/NEJMoa1301296.
[23] YAMANOI K, BABA T, ABIKO K, et al. Acquisition of a side population fraction augments malignant phenotype in ovarian cancer [J]. Sci Rep, 2019, 9 (1): 14215. DOI: 10.1038/s41598-019-50794-w.
[24] ZHANG H, HUANG C J, TIAN Y, et al. Ectopic overexpression of COTE1 promotes cellular invasion of hepatocellular carcinoma [J]. Asian Pac J Cancer Prev, 2012, 13 (11): 5799-5804. DOI: 10.7314/apjcp.2012.13.11.5799.
[25] WU C L, TAN Q X, LIU D, et al. High FAM189B expression and its prognostic value in patients with gastric cancer [J]. Biomed Res Int, 2021, 2021: 8875971. DOI: 10.1155/2021/8875971.
[26] ZHANG G H, SU L S, LV X P, et al. A novel tumor doubling time-related immune gene signature for prognosis prediction in hepatocellular carcinoma [J]. Cancer Cell Int, 2021, 21 (1): 522. DOI: 10.1186/s12935-021-02227-w.
[27] JANG H, JUN Y, KIM S, et al. FCN3 functions as a tumor suppressor of lung adenocarcinoma through induction of endoplasmic reticulum stress [J]. Cell Death Dis, 2021, 12 (4): 407. DOI: 10.1038/s41419-021-03675-y.

（收稿日期：2023-03-17；修回日期：2023-08-18）
（本文编辑：贾萌萌）